(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2024년02월08일
(11) 등록번호 10-2635611
(24) 등록일자 2024년02월06일

(51) 국제특허분류(Int. Cl.)
G06T 9/00 (2019.01) *G06N 3/04* (2023.01)
G06N 3/08 (2023.01) *G06N 5/00* (2023.01)
H04N 19/124 (2014.01) *H04N 19/13* (2014.01)
H04N 19/184 (2014.01)
(52) CPC특허분류
G06T 9/002 (2013.01)
G06N 3/045 (2023.01)
(21) 출원번호 10-2022-0101413
(22) 출원일자 2022년08월12일
심사청구일자 2022년08월12일
(56) 선행기술조사문헌
CN114330554 A*
David Minnen 외 2명, "Joint Autoregressive and Hierarchical Priors for Learned Image Compression", Advances in Neural Information Processing Systems 31(NeurIPS 2018), 페이지 1-10, 2018년 공개. 1부.*
Yan Qu 외 2명, "H-AT: Hybrid Attention Transfer for Knowledge Distillation", Pattern Recognition and Computer Vision, Third Chinese Conference, PRCV 2020, Nanjing, China, 페이지 249-259, 2020.10.13. 1부.*
Zhilu Zhang 외 5명, "Learning RAW-to-sRGB Mappings with Inaccurately Aligned Supervision", Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 페이지 4348-4358, 2021년 공개. 1부.*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
정승원
서울특별시 서초구 서초대로65길 13-10, 113동 1603호(서초동, 서초래미안아파트)
정우석
서울특별시 서대문구 서소문로 45, 423호(합동)
(74) 대리인
송인호, 윤형근, 최관락

전체 청구항 수 : 총 5 항 심사관 : 권영학

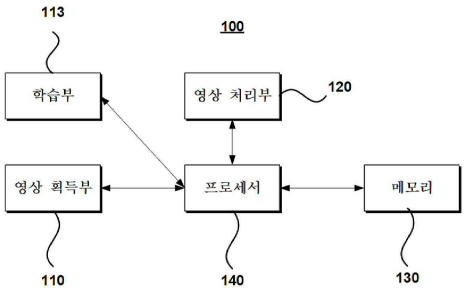
(54) 발명의 명칭 **영상 신호 처리 장치 및 이의 영상 처리 방법**

(57) 요 약

영상 신호 처리 장치 및 이의 영상 처리 방법이 개시된다. 영상 신호 처리 장치는, 원시 데이터를 획득하는 영상 입력부; 및 상기 원시 데이터를 학습된 딥러닝 기반 압축스트림 생성 모델에 적용하여 손실 압축 영상 생성이 가능한 비트스트림을 생성하고, 상기 비트스트림을 이용하여 손실 압축 영상을 복원하는 영상 처리부를 포함하되,

(뒷면에 계속)

대 표 도 - 도1



상기 딥러닝 기반 압축스트림 생성 모델은 서로 상이한 복수의 교사 네트워크 모델을 통해 지식 증류받을 수 있다.

(52) CPC특허분류
G06N 3/082 (2023.01)
G06N 5/00 (2022.01)
H04N 19/124 (2015.01)
H04N 19/13 (2015.01)
H04N 19/184 (2015.01)
G06T 2207/20084 (2013.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711157106
과제번호 2022R1A2C2002810
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 개인기초연구(과기정통부)
연구과제명 Light-to-bit 방식의 종단종 SPAD 카메라 ISP 네트워크 연구
기 여 율 1/1
과제수행기관명 고려대학교 산학협력단
연구기간 2022.03.01 ~ 2023.02.28
공지예외적용 : 있음

명 세 서

청구범위

청구항 1

제1 교사 네트워크 모델과 제2 교사 네트워크 모델을 가지며, 제1 트레이닝 데이터를 이용하여 sRGB 영상을 인코딩하고 복원하도록 상기 제1 교사 네트워크 모델을 학습하고, 제2 트레이닝 데이터를 이용하여 원시 데이터를 비트스트림으로 인코딩하고 복원하도록 제2 교사 네트워크 모델을 학습시키는 학습부;

상기 원시 데이터를 획득하는 영상 입력부; 및

상기 원시 데이터를 학습된 딥러닝 기반 압축스트림 생성 모델에 적용하여 손실 압축 영상 생성이 가능한 비트스트림을 생성하고, 상기 비트스트림을 이용하여 손실 압축 영상을 복원하는 영상 처리부를 포함하되,

상기 딥러닝 기반 압축스트림 생성 모델은,

4개의 컨볼루션 레이어를 가지되, 상기 4개의 컨볼루션 레이어를 통해 상기 원시 데이터를 인코딩하여 손실 압축 영상에 대한 비트스트림을 생성하는 인코더; 및

5개의 디컨볼루션 레이어를 가지며, 상기 5개의 디컨볼루션 레이어를 통해 상기 비트스트림을 재구성하여 손실 압축 영상을 복원하는 디코더를 포함하되,

상기 인코더와 상기 디코더는 서로 상이한 복수의 교사 네트워크 모델을 통해 지식 증류받되,

상기 제1 및 제2 교사 네트워크 모델은 각각 인코더와 디코더가 동일한 개수의 중간 레이어를 가지는 대칭 구조로 구성되되,

상기 딥러닝 기반 압축스트림 생성 모델의 인코더와 디코더는 중간 레이어의 개수가 상이한 비대칭 구조이되, 상기 딥러닝 기반 압축스트림 생성 모델의 인코더와 상기 제2 교사 네트워크 모델의 인코더는 동일 구조로 구성되며, 상기 딥러닝 기반 압축스트림 생성 모델의 디코더와 상기 제1 교사 네트워크 모델의 디코더는 동일 구조로 구성되되,

상기 학습된 제2 교사 네트워크 모델의 인코더는 상기 딥러닝 기반 압축스트림 생성 모델의 인코더로 지식증류되며, 상기 학습된 제1 교사 네트워크 모델의 디코더는 상기 딥러닝 기반 압축스트림 생성 모델의 디코더로 지식증류되는 것을 특징으로 하는 영상 신호 처리 장치.

청구항 2

삭제

청구항 3

삭제

청구항 4

삭제

청구항 5

제1 항에 있어서,

상기 제1 트레이닝 데이터는 sRGB 영상이고, 상기 제2 트레이닝 데이터는 원시 데이터인 것을 특징으로 하는 영상 신호 처리 장치.

청구항 6

제1 항에 있어서,

상기 제2 교사 네트워크 모델의 인코더와 상기 딥러닝 기반 압축스트림 생성 모델의 인코더의 각 컨볼루션 레이어에서 출력되는 어텐션 맵 사이의 평균 제곱 오차가 최소가 되도록 지식 증류되며,

상기 제1 교사 네트워크 모델의 디코더의 디컨볼루션 레이어의 출력과 와 상기 딥러닝 기반 압축스트림 생성 모델의 디코더의 디컨볼루션 레이어의 어텐션 맵 사이의 평균 제곱 오차가 최소가 되도록 지식 증류되는 것을 특징으로 하는 영상 신호 처리 장치.

청구항 7

(a) 제1 트레이닝 데이터를 이용하여 제1 교사 네트워크 모델을 학습하고, 제2 트레이닝 데이터를 이용하여 제2 교사 네트워크 모델을 학습하는 단계-제1 트레이닝 데이터를 이용하여 sRGB 영상을 인코딩하고 복원하도록 상기 제1 교사 네트워크 모델을 학습하고, 제2 트레이닝 데이터를 이용하여 원시 데이터를 비트스트림으로 인코딩하고 복원하도록 제2 교사 네트워크 모델을 학습함;

(b) 상기 제1 교사 네트워크 모델의 일부와 상기 제2 교사 네트워크 모델의 일부를 딥러닝 기반 압축스트림 생성 모델에 지식 증류하는 단계; 및

(c) 영상 센서에서 촬영된 원시 데이터를 상기 딥러닝 기반 압축스트림 생성 모델에 적용하여 손실 압축 영상 생성이 가능한 비트스트림을 생성하고, 상기 비트스트림을 이용하여 손실 압축 영상을 복원하는 단계를 포함하되,

상기 딥러닝 기반 압축스트림 생성 모델은,

4개의 컨볼루션 레이어를 가지되, 상기 4개의 컨볼루션 레이어를 통해 상기 원시 데이터를 인코딩하여 손실 압축 영상에 대한 비트스트림을 생성하는 인코더; 및

5개의 디컨볼루션 레이어를 가지며, 상기 5개의 디컨볼루션 레이어를 통해 상기 비트스트림을 재구성하여 손실 압축 영상을 복원하는 디코더를 포함하되,

상기 인코더와 상기 디코더는 서로 상이한 복수의 교사 네트워크 모델을 통해 지식 증류받되,

상기 제1 및 제2 교사 네트워크 모델은 각각 인코더와 디코더가 동일한 개수의 중간 레이어를 가지는 대칭 구조로 구성되되,

상기 딥러닝 기반 압축스트림 생성 모델의 인코더와 디코더는 중간 레이어의 개수가 상이한 비대칭 구조이되, 상기 딥러닝 기반 압축스트림 생성 모델의 인코더와 상기 제2 교사 네트워크 모델의 인코더는 동일 구조로 구성되며, 상기 딥러닝 기반 압축스트림 생성 모델의 디코더와 상기 제1 교사 네트워크 모델의 디코더는 동일 구조로 구성되되,

상기 학습된 제2 교사 네트워크 모델의 인코더는 상기 딥러닝 기반 압축스트림 생성 모델의 인코더로 지식증류되며, 상기 학습된 제1 교사 네트워크 모델의 디코더는 상기 딥러닝 기반 압축스트림 생성 모델의 디코더로 지식증류되는 것을 특징으로 하는 영상 신호 처리 장치의 영상 처리 방법.

청구항 8

제7 항에 있어서,

상기 (b) 단계는,

상기 제2 교사 네트워크 모델의 인코더와 상기 딥러닝 기반 압축스트림 생성 모델의 인코더의 각 컨볼루션 레이어에서 출력되는 어텐션 맵 사이의 평균 제곱 오차가 최소가 되도록 상기 제2 교사 네트워크 모델의 인코더를 상기 딥러닝 기반 압축스트림 생성 모델의 인코더로 지식 증류하고,

상기 제1 교사 네트워크 모델의 디코더의 디컨볼루션 레이어의 출력과 상기 딥러닝 기반 압축스트림 생성 모델

의 디코더의 디컨볼루션 레이어의 어텐션 맵 사이의 평균 제곱 오차가 최소가 되도록 상기 제1 교사 네트워크 모델의 디코더를 상기 딥러닝 기반 압축스트림 생성 모델의 디코더로 지식 증류하는 것을 특징을 하는 영상 신호 처리 장치의 영상 처리 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 영상 신호 처리 장치 및 이의 영상 처리 방법에 관한 것이다.

배 경 기 술

[0003] 스마트폰의 처리 능력이 향상됨에 따라 작은 영상 센서로 인한 제약을 극복하기 위해 딥러닝을 이용한 영상 신호 처리에 대해 많은 연구가 이루어지고 있다. 영상 센서의 해상도가 증가함에 따라 생성된 sRGB 영상의 용량도 증가하게 되었으며, 이러한 영상의 저장 및 전송을 위해서는 손실 압축이 필수적이다. 손실 압축을 고려한 영상 신호 처리에 대한 연구가 일부 수행된바 있으나 비트스트림을 생성하는 기술은 없는 상황이다.

발명의 내용

해결하려는 과제

[0005] 본 발명은 영상 신호 처리 장치 및 이의 영상 처리 방법을 제공하기 위한 것이다.

[0006] 또한, 본 발명은 영상 처리 및 손실 압축을 통합하여 영상 센서의 로우 데이터로부터 손실 압축 영상 생성이 가능한 비트스트림을 출력할 수 있는 영상 신호 처리 장치 및 이의 영상 처리 방법을 제공하기 위한 것이다.

[0007] 이를 통해, 본 발명은 영상 데이터의 용량을 줄일 수 있는 영상 신호 처리 장치 및 이의 영상 처리 방법을 제공하기 위한 것이다.

[0008] 또한, 본 발명은 영상 처리와 손실 압축에 특화된 네트워크를 이용하여 지식증류를 수행하여 성능이 향상된 영상 신호 처리 장치 및 이의 영상 처리 방법을 제공하기 위한 것이다.

과제의 해결 수단

[0010] 본 발명의 일 측면에 따르면 원시 데이터에 대한 영상 처리 및 손실 압축이 가능한 영상 신호 처리 장치가 제공된다.

[0011] 본 발명의 일 실시예에 따르면, 원시 데이터를 획득하는 영상 입력부; 및 상기 원시 데이터를 학습된 딥러닝 기반 압축스트림 생성 모델에 적용하여 손실 압축 영상 생성이 가능한 비트스트림을 생성하고, 상기 비트스트림을 이용하여 손실 압축 영상을 복원하는 영상 처리부를 포함하되, 상기 딥러닝 기반 압축스트림 생성 모델은 서로 상이한 복수의 교사 네트워크 모델을 통해 지식 증류받는 것을 특징으로 하는 영상 신호 처리 장치가 제공될 수 있다.

[0012] 상기 딥러닝 기반 압축스트림 생성 모델은, 4개의 컨볼루션 레이어를 가지되, 상기 4개의 컨볼루션 레이어를 통해 상기 원시 데이터를 인코딩하여 손실 압축 영상에 대한 비트스트림을 생성하는 인코더; 및 5개의 디컨볼루션 레이어를 가지며, 상기 5개의 디컨볼루션 레이어를 통해 상기 비트스트림을 재구성하여 손실 압축 영상을 복원하는 디코더를 포함하되, 상기 인코더와 상기 디코더는 서로 상이한 복수의 교사 네트워크 모델을 통해 지식 증류받을 수 있다.

[0013] 제1 교사 네트워크 모델과 제2 교사 네트워크 모델을 가지며, 제1 트레이닝 데이터를 이용하여 sRGB 영상을 인코딩하고 복원하도록 상기 제1 교사 네트워크 모델을 학습하고, 제2 트레이닝 데이터를 이용하여 원시 데이터를 비트스트림으로 인코딩하고 복원하도록 제2 교사 네트워크 모델을 학습시키는 학습부를 더 포함할 수 있다.

[0014] 상기 제1 및 제2 교사 네트워크 모델은 각각 인코더와 디코더가 동일한 개수의 중간 레이어를 가지는 대칭 구조로 구성되되, 상기 딥러닝 기반 압축스트림 생성 모델의 인코더와 디코더는 중간 레이어의 개수가 상이한 비대칭 구조이되, 상기 딥러닝 기반 압축스트림 생성 모델의 인코더와 상기 제2 교사 네트워크 모델의 인코더는 동일 구조로 구성되며, 상기 딥러닝 기반 압축스트림 생성 모델의 디코더와 상기 제1 교사 네트워크 모델의 디코

더는 동일 구조로 구성되되, 상기 학습된 제2 교사 네트워크 모델의 인코더는 상기 딥러닝 기반 압축스트림 생성 모델의 인코더로 지식증류되며, 상기 학습된 제1 교사 네트워크 모델의 디코더는 상기 딥러닝 기반 압축스트림 생성 모델의 인코더로 지식증류될 수 있다.

[0015] 상기 제1 트레이닝 데이터는 sRGB 영상이고, 상기 제2 트레이닝 데이터는 원시 데이터이다.

[0016] 상기 제2 교사 네트워크 모델의 인코더와 상기 딥러닝 기반 압축스트림 생성 모델의 인코더의 각 컨볼루션 레이어에서 출력되는 어텐션 맵 사이의 평균 제곱 오차가 최소가 되도록 지식 증류되며, 상기 제1 교사 네트워크 모델의 디코더의 디컨볼루션 레이어의 출력과 와 상기 딥러닝 기반 압축스트림 생성 모델의 디코더의 디컨볼루션 레이어의 어텐션 맵 사이의 평균 제곱 오차가 최소가 되도록 지식 증류될 수 있다.

[0018] 본 발명의 다른 측면에 따르면, 지식 증류에 기초한 원시 데이터에 대한 영상 처리 및 손실 압축이 가능한 영상 신호 처리 장치의 영상 처리 방법이 제공된다.

[0019] 본 발명의 일 실시예에 따르면, (a) 제1 트레이닝 데이터를 이용하여 제1 교사 네트워크 모델을 학습하고, 제2 트레이닝 데이터를 이용하여 제2 교사 네트워크 모델을 학습하는 단계; (b) 상기 제1 교사 네트워크 모델의 일부와 상기 제2 교사 네트워크 모델의 일부를 딥러닝 기반 압축스트림 생성 모델에 지식 증류하는 단계; 및 (c) 영상 센서에서 촬영된 원시 데이터를 상기 딥러닝 기반 압축스트림 생성 모델에 적용하여 손실 압축 영상 생성이 가능한 비트스트림을 생성하고, 상기 비트스트림을 이용하여 손실 압축 영상을 복원하는 단계를 포함하는 영상 신호 처리 장치의 영상 처리 방법이 제공될 수 있다.

[0020] 상기 (b) 단계는, 상기 제2 교사 네트워크 모델의 인코더와 상기 딥러닝 기반 압축스트림 생성 모델의 인코더의 각 컨볼루션 레이어에서 출력되는 어텐션 맵 사이의 평균 제곱 오차가 최소가 되도록 상기 제2 교사 네트워크 모델의 인코더를 상기 딥러닝 기반 압축스트림 생성 모델의 인코더로 지식 증류하고, 상기 제1 교사 네트워크 모델의 디코더의 디컨볼루션 레이어의 출력과 상기 딥러닝 기반 압축스트림 생성 모델의 디코더의 디컨볼루션 레이어의 어텐션 맵 사이의 평균 제곱 오차가 최소가 되도록 상기 제1 교사 네트워크 모델의 디코더를 상기 딥러닝 기반 압축스트림 생성 모델의 디코더로 지식 증류할 수 있다.

발명의 효과

[0022] 본 발명의 일 실시예에 영상 신호 처리 장치 및 이의 영상 처리 방법을 제공함으로써, 영상 처리 및 손실 압축을 통합하여 영상 센서의 로우 데이터로부터 손실 압축 영상 생성이 가능한 비트스트림을 출력할 수 있다.

[0023] 이를 통해, 본 발명은 영상 데이터의 용량을 줄일 수 있는 이점이 있다.

[0024] 또한, 본 발명은 영상 처리와 손실 압축에 특화된 네트워크를 이용하여 지식증류를 수행하여 성능을 향상시킬 수 있는 이점도 있다.

도면의 간단한 설명

[0026] 도 1은 본 발명의 일 실시예에 따른 영상 신호 처리 장치의 내부 구성을 개략적으로 도시한 도면.

도 2는 본 발명의 일 실시예에 따른 증류 모델을 개략적으로 도시한 도면.

도 3은 본 발명의 일 실시예에 따른 두 교사 네트워크 모델을 이용한 지식 증류를 설명하기 위해 도시한 도면.

도 4는 본 발명의 일 실시예에 따른 영상 신호 처리 장치의 영상 처리 방법을 나타낸 순서도.

도 5는 종래와 본 발명의 일 실시예에 따른 원시 데이터와 재구성된 sRGB 영상간의 평균 PSNR과 픽셀당 평균 비트수(bpp)를 비교한 결과를 도시한 도면.

도 6은 종래와 본 발명의 일 실시예에 따른 실제 sRGB 영상과 복원된 sRGB 영상간의 에러 맵을 시각화한 결과를 도시한 도면.

도 7은 종래와 본 발명의 일 실시예에 따른 복원 결과를 도시한 도면.

발명을 실시하기 위한 구체적인 내용

[0027] 본 명세서에서 사용되는 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "구성된다" 또는 "포함한다" 등의 용어는 명세서상에 기재된 여러 구성 요소들, 또는 여러 단계들을 반드시 모두 포함하는 것으로 해석되지 않아야 하며, 그 중 일부 구성 요소들 또는 일부 단계들은 포함되지

않을 수도 있고, 또는 추가적인 구성 요소 또는 단계들을 더 포함할 수 있는 것으로 해석되어야 한다. 또한, 명세서에 기재된 "...부", "모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다.

[0028] 이하, 첨부된 도면들을 참조하여 본 발명의 실시예를 상세히 설명한다.

[0030] 도 1은 본 발명의 일 실시예에 따른 영상 신호 처리 장치의 내부 구성을 개략적으로 도시한 도면이고, 도 2는 본 발명의 일 실시예에 따른 증류 모델을 개략적으로 도시한 도면이고, 도 3은 본 발명의 일 실시예에 따른 두 교사 네트워크 모델을 이용한 지식 증류를 설명하기 위해 도시한 도면이다.

[0031] 도 1을 참조하면, 본 발명의 일 실시예에 따른 영상 신호 처리 장치(100)는 영상 획득부(110), 학습부(113), 영상 처리부(120), 메모리(130) 및 프로세서(140)를 포함하여 구성된다.

[0032] 영상 획득부(110)는 센서(미도시)로부터 피사체를 촬영한 원시 데이터를 획득하기 위한 수단이다.

[0033] 여기서, 원시 데이터는 로우 데이터(RAW data)일 수 있다.

[0034] 학습부(113)는 제1 트레이닝 데이터 및 제2 트레이닝 데이터를 이용하여 제1 교사 네트워크 모델 및 제2 교사 네트워크 모델을 각각 학습하기 위한 수단이다. 여기서, 제1 교사 네트워크 모델 및 제2 교사 네트워크 모델은 각각 인코더와 디코더로 구성되는 오토인코더 모델을 기반으로 할 수 있다. 또한, 도 3에서 보여지는 바와 같이, 각각 인코더와 디코더가 동일한 개수의 중간 레이어를 가지는 대칭 구조로 구성될 수 있다.

[0035] 제1 트레이닝 데이터는 sRGB 영상일 수 있으며, 제2 트레이닝 데이터는 원시 데이터일 수 있다.

[0036] 또한, 제1 교사 네트워크 모델과 제2 교사 네트워크 모델은 각각 인코더와 디코더 모델일 수 있다. 제1 네트워크 교사 모델은 제1 트레이닝 데이터를 이용하여 sRGB 영상을 인코딩하고 복원하도록 학습될 수 있다. 또한, 제2 교사 네트워크 모델은 제2 트레이닝 데이터를 이용하여 원시 데이터를 인코딩하여 비트스트림을 생성하고, 해당 비트스트림을 이용하여 원시 데이터를 복원하도록 학습될 수 있다.

[0037] 즉, 제1 교사 네트워크 모델은 sRGB 영상 처리에 특화되도록 학습될 수 있으며, 제2 교사 네트워크 모델은 원시 데이터를 압축하는데 특화되도록 학습될 수 있다.

[0038] 학습이 완료된 이후, 제1 교사 네트워크 모델과 제2 교사 네트워크 모델의 일부는 각각 하기에서 설명되는 압축 스트림 생성 모델에 지식 증류될 수 있다.

[0039] 영상 처리부(120)는 영상 획득부(110)에서 획득된 원시 데이터를 딥러닝 기반 압축스트림 생성 모델에 적용하여 손실 압축 영상 생성이 가능한 비트스트림을 생성하고, 해당 비트스트림을 이용하여 손실 압축 영상을 복원하기 위한 장치이다.

[0040] 예를 들어, 영상 처리부(120)는 원시 데이터(RAW)를 딥러닝 기반 압축스트림 생성 모델에 적용하여 sRGB 영상 생성이 가능한 비트스트림을 생성하고, 해당 비트스트림을 이용하여 sRGB 영상을 복원할 수 있다.

[0041] 이에 대해 도 2를 참조하여 보다 상세히 설명하기로 한다.

[0042] 도 2에는 딥러닝 기반 압축스트림 생성 모델이 도시되어 있다. 도 2에 도시된 바와 같이, 딥러닝 기반 압축스트림 생성 모델은 인코더(210)와 디코더(220)로 구성된다.

[0043] 인코더(210)는 일련의 스트라이드 컨볼루션 레이어와 GDN(Generalized divisive normalization)를 포함하며, 원시 데이터를 입력받아 3채널 sRGB 영상 압축을 위한 비트스트림을 생성할 수 있다.

[0044] 인코더(210)는 4개 채널로 압축된 원시 데이터를 사용하기 때문에 큰 커널 크기를 사용하면 수용 필드가 제어할 수 없을 정도로 빠르게 커질 수 있다. 이로 인해, 채널 수를 256으로 늘리면서 커널 크기 3을 사용할 수 있다. 또한, 마지막 컨볼루션 레이어는 192개 채널로 잠재 표현을 생성할 수 있다.

[0045] 디코더(220)는 5개 디컨볼루션 레이어를 가지며, RCAG(residual channel Attention Group)에서 두 개의 잔여 채널 주의 블록이 사용될 수 있다. 디코더(220)에서 디코딩된 잠재 벡터만 손실 압축 영상(예를 들어, sRGB)을 재구성할 수 있어야 하므로 인코더(210)와 디코더(220) 사이에는 스킵(skip) 연결이 존재하지 않는다.

[0046] 도 3에서 보여지는 바와 같이, 본 발명의 일 실시예에 따른 증류 모델인 딥러닝 기반 압축스트림 생성 모델의 인코더(210)와 디코더(220)는 중간 레이어의 개수가 동일하지 않은 비대칭 구조로 구성될 수 있다.

[0047] 이러한, 딥러닝 기반 압축스트림 생성 모델은 인코더(210)와 디코더(220)가 서로 다른 교사 네트워크 모델을 통

해 지식 증류받을 수 있다.

[0048] 도 3에 도시된 바와 같이, 본 발명의 일 실시예에 따르면, 딥러닝 기반 압축스트림 생성 모델의 인코더(210)는 제2 교사 네트워크 모델에 의해 지식 증류될 수 있다. 여기서, 제2 교사 네트워크 모델의 인코더(310b)는 원시 데이터($x_r \in \mathbb{R}^{4\times H/2\times W/2}$)를 입력으로 취하고, 원시 데이터 ($\hat{x}_r \in \mathbb{R}^{4\times H/2\times W/2}$)를 재구성하기 위한 비트스트림을 생성하도록 학습될 수 있다.

[0049] 제2 교사 네트워크 모델의 인코더는 효율적인 압축을 위해 간결한 표현을 찾도록 훈련되었기 때문에 제2 교사 네트워크 모델의 인코더의 지식을 딥러닝 기반 압축스트림 생성 모델의 인코더(210)로 증류시킴으로써 당해 딥러닝 기반 압축스트림 생성 모델의 인코더(210)의 성능을 높일 수 있는 이점이 있다,

[0050] 제2 교사 네트워크 모델의 인코더(330a)와 디코더(340b)는 대칭 구조로 구성되며, 동일한 개수의 중간 레이어를 가지도록 구성될 수 있다. 즉, 인코더(330a)가 4개의 컨볼루션 레이어를 가지는 경우, 디코더(340b) 또한 4개의 디컨볼루션 레이어를 가지도록 구성될 수 있다.

[0051] 이러한, 제2 교사 네트워크 모델의 인코더(330a)는 원시 데이터를 4개의 컨볼루션 레이어를 통과시켜 손실 압축 영상 복원이 가능한 비트스트림을 생성하도록 학습되며, 디코더(340b)는 해당 비트스트림을 이용하여 원시 데이터를 복원하도록 학습될 수 있다.

[0052] 제1 교사 네트워크 모델은 잠재 표현에서 sRGB 영상을 재구성하는데 집중하도록 디코더(320b)의 지식을 딥러닝 기반 압축스트림 생성 모델의 디코더(220)로 증류시킬 수 있다.

[0053] 제1 교사 네트워크 모델 또한 인코더(310a)와 디코더(320b)가 동일한 개수의 중간 레이어를 가지는 대칭 구조로 구성될 수 있다.

[0054] 두 교사 네트워크 모델을 통해 지식 증류가 용이하도록 딥러닝 기반 압축스트림 생성 모델은 인코더(210)와 디코더(220)가 비대칭 구조로 구성되되, 딥러닝 기반 압축스트림 생성 모델의 인코더(210)는 제2 교사 네트워크 모델의 인코더(330a)와 동일한 구조로 구성되며, 딥러닝 기반 압축스트림 생성 모델의 디코더(220)는 제1 교사 네트워크 모델의 디코더(320b)와 동일한 구조로 구성될 수 있다.

[0055] 이를 통해, 제1 및 제2 교사 네트워크 모델의 각 일부와 딥러닝 기반 압축 스트림 생성 모델의 중간 레이어에서 생성된 공간 어텐션 맵을 비교하는 어텐션 전달이 용이하도록 할 수 있다.

[0056] 공간 어텐션 맵은 중간 레이어 출력의 채널 차원에 따른 절대값의 합으로 정의될 수 있다. 그러나, 절대값을 취하는 것이 비효율적이므로, 본 발명의 일 실시예에서는 채널 차원에 따른 직접 합으로 어텐션 맵을 정의하기로 한다.

[0057] 이를 수학식으로 나타내면, 수학식 1과 같다.

[0058] [수학식 1]

$$M_j = \sum_{i=1}^{C_j} A_j\,(i,:,:),$$

[0059]

[0060] 여기서, $A_j \in \mathbb{R}^{C_j\times H_j\times W_j}$ 는 j번째 중간 레이어의 출력을 나타내고,

M_j는 j번째 중간 레이어의 어텐션 맵을 나타내며, C_j, H_j, W_j는 각각 채널 차원, 높이 및 너비를 나타낸다.

[0061] 컨볼루션 레이어 출력(ReLU 활성화 전에) 사이에 어텐션 전달이 적용되기 때문에, 절대값을 취하면 출력 텐서(output tensor)에 포함된 방향 정보가 손실될 수 있다. 지식 증류의 어텐션 손실, L_{AT},은 교사 네트워크 모델과 딥러닝 기반 압축스트림 생성 모델 사이의 정규화된 어텐션 맵 사이의 평균 제곱 오차로 정의될 수 있다. 이를 수학식으로 나타내면, 수학식 2와 같이 나타낼 수 있다.

[0062] [수학식 2]

$$L_{AT} = \sum_{j=1}^{n_p} \alpha_j L_{AT}^j,$$

$$L_{AT}^j = \frac{1}{N_j} \left\| \frac{M_j^S}{\|M_j^S\|_2} - \frac{M_j^T}{\|M_j^T\|_2} \right\|_2^2$$

[0063]

[0064] 여기서, $\|\cdot\|_2$는 L2-norm 함수를 나타내고, $M_j^S(M_j^T)$는 딥러닝 기반 압축스트림 생성 모델(교사 네트워크 모델)의 j번째 어텐션 맵을 나타내며, n_p는 어텐션 맵의 쌍의 개수를 나타내고, N_j는 j 어텐션 맵의 요소 개수를 나타내며, α_j는 j번째 손실 항의 가중치를 나타낸다. 초기 훈련 과정 동안 율-왜곡 손실 보다 상대적으로 어텐션 손실이 높게 설정되고, 훈련이 진행됨에 따라 감쇠되도록 α_j가 설정될 수 있다.

[0065] 이와 같은 방식으로 딥러닝 기반 압축스트림 생성 모델은 처음에는 지식 증류에 초점을 맞추고 점차적으로 비율-왜곡 최적화를 주요 목표로 전화될 수 있다.

[0066] 이를 위해 α_j는 수학식 3과 같이 선택될 수 있다.

[0067] [수학식 3]

$$\alpha_j = \alpha_0 \cdot \gamma^k$$

[0068]

[0069] 여기서, α_0 는 초기값을 나타내고, γ 는 감쇠 계수를 나타낸다. α_0 는 두 인코더 사이의 지식 증류를 위해 10^6 으로 설정되고, 두 디코더 사이의 지식 증류를 위해 10^5 로 설정될 수 있다. 또한, γ 는 인코더와 디코더 두 경우 모두 0.999992로 설정될 수 있다. 훈련 횟수 k가 증가할수록 α_j 는 천천히 감소될 수 있다.

[0070] 최종 손실 함수는 수학식 4와 같이 정의될 수 있다.

[0071] [수학식 4]

$$L_{total} = L_R + \lambda L_D + L_{AT}$$

[0072]

[0073] 여기서, L_R 과 L_D 는 정의된 비율 및 왜곡 손실 항을 나타내고, Lagrangian 승수 λ 는 비율-왜곡 트레이드 오프를 결정할 수 있다.

[0074] 다시 정리하면, 딥러닝 기반 압축스트림 생성 모델의 인코더와 디코더는 서로 다른 교사 네트워크를 통해 지식 증류 받을 수 있다.

[0075] 상술한 바와 같이, 제2 교사 네트워크 모델과 제1 교사 네트워크 모델을 압축 및 영상 처리에 특화되도록 사전 학습시킨 후 증류 모델인 딥러닝 기반 압축스트림 생성 모델의 인코더와 디코더로 지식 증류함으로써 성능을 더 향상시킬 수 있는 이점이 있다.

[0076] 메모리(130)는 본 발명의 일 실시예에 따른 두 교사 네트워크 모델의 지식 증류를 이용한 딥러닝 기반 압축스트림 생성 모델에서 손실 압축 영상 생성이 가능한 비트스트림 생성 및 복원을 위한 영상 처리 방법을 수행하기 위한 프로그램 코드를 저장한다.

[0077] 프로세서(140)는 본 발명의 일 실시예에 따른 영상 신호 처리 장치(100)의 내부 구성 요소들(예를 들어, 영상 획득부(110), 영상 처리부(120), 메모리(130) 등)을 제어하기 위한 수단이다.

[0079] 도 4는 본 발명의 일 실시예에 따른 영상 신호 처리 장치의 영상 처리 방법을 나타낸 순서도이고, 도 5는 종래와 본 발명의 일 실시예에 따른 원시 데이터와 재구성된 sRGB 영상간의 평균 PSNR과 픽셀당 평균 비트수(bpp)를 비교한 결과를 도시한 도면이며, 도 6은 종래와 본 발명의 일 실시예에 따른 실제 sRGB 영상과 복원된 sRGB 영상간의 에러 맵을 시각화한 결과를 도시한 도면이고, 도 7은 종래와 본 발명의 일 실시예에 따른 복원 결과를

도시한 도면이다.

[0080] 단계 410에서 영상 신호 처리 장치(100)는 제1 교사 네트워크 모델과 제2 교사 네트워크 모델을 학습시킨다.

[0081] 예를 들어, 영상 신호 처리 장치(100)는 제1 트레이닝 데이터를 이용하여 제1 교사 네트워크 모델을 학습시키고, 제2 트레이닝 데이터를 이용하여 제2 교사 네트워크 모델을 학습시킨다.

[0082] 여기서, 제1 트레이닝 데이터와 제2 트레이닝 데이터는 서로 다른 유형(포맷)의 데이터이다. 전술한 바와 같이, 제1 트레이닝 데이터는 sRGB 영상이고, 제2 트레이닝 데이터는 RAW 데이터일 수 있다.

[0083] 제1 교사 네트워크 모델은 sRGB 영상을 인코딩하고, 그 결과를 이용하여 다시 sRGB 영상을 복원하도록 학습될 수 있다.

[0084] 또한, 제2 교사 네트워크 모델은 RAW 데이터를 인코딩하여 비트스트림을 생성하고, 해당 비트스트림을 이용하여 RAW 데이터를 복원하도록 학습될 수 있다.

[0085] 단계 415에서 영상 신호 처리 장치(100)는 제1 교사 네트워크 모델과 제2 교사 네트워크 모델의 일부를 각각 딥러닝 기반 압축스트림 생성 모델에 지식증류한다.

[0086] 이미 전술한 바와 같이, 제1 교사 네트워크 모델의 디코더는 딥러닝 기반 압축스트림 생성 모델의 디코더로 지식 증류되며, 제2 교사 네트워크 모델의 인코더는 딥러닝 기반 압축스트림 생성 모델의 인코더로 지식 증류될 수 있다.

[0087] 도 2 및 도 3을 참조하여 설명한 바와 같이, 지식 증류함에 있어, 지식 증류 손실은 각 교사 네트워크의 중간 레이어의 결과값(예를 들어, 어텐션 맵)과 딥러닝 기반 압축스트림 생성 모델의 중간 레이어의 결과값 사이의 최소 제곱 오차가 최소가 되는 것을 목표로 설계될 수 있다.

[0088] 단계 420에서 영상 신호 처리 장치(100)는 영상 센서에서 획득된 원시 데이터를 딥러닝 기반 압축스트림 생성 모델에 적용하여 손실 압축 영상 생성이 가능한 비트스트림을 생성하고, 해당 비트스트림을 이용하여 손실 압축 영상을 복원한다.

[0089] 단계 425에서 영상 신호 처리 장치(100)는 비트스트림을 외부로 출력한다.

[0090] 도 5는 종래와 본 발명의 일 실시예에 따른 원시 데이터와 재구성된 sRGB 영상간의 평균 PSNR과 픽셀당 평균 비트수(bpp)를 비교한 결과를 도시한 도면이다.

[0091] 도 5에서 보여지는 바와 같이, 본 발명의 일 실시예에 따른 교사 네트워크 모델을 통해 지식 증류받은 딥러닝 기반 압축스트림 생성 모델의 성능이 종래에 비해 향상된 것을 알 수 있다.

[0092] 도 6은 종래와 본 발명의 일 실시예에 따른 실제 sRGB 영상과 복원된 sRGB 영상간의 에러 맵을 시각화한 결과를 도시한 도면이다. 도 6에서 보여지는 바와 같이, 본 발명의 일 실시예에 따른 두 교사 네트워크 모델을 통해 지식 증류받은 딥러닝 기반 압축스트림 생성 모델이 텍스처가 풍부한 영역에서 왜곡을 줄여 유사한 압축률에서 최고 영상 품질을 제공함을 알 수 있다.

[0093] 도 7은 종래와 본 발명의 일 실시예에 따른 복원 결과를 도시한 도면이다. 도 7에서 보여지는 바와 같이, 종래 방법에 비해 본 발명의 일 실시예에 따른 두 교사 네트워크 모델을 통해 지식 증류받은 딥러닝 기반 압축스트림 생성 모델이 색상을 더 정확하게 렌더링하는 것을 알 수 있다.

[0095] 본 발명의 실시 예에 따른 장치 및 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 컴퓨터 판독 가능 매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 분야 통상의 기술자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media) 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

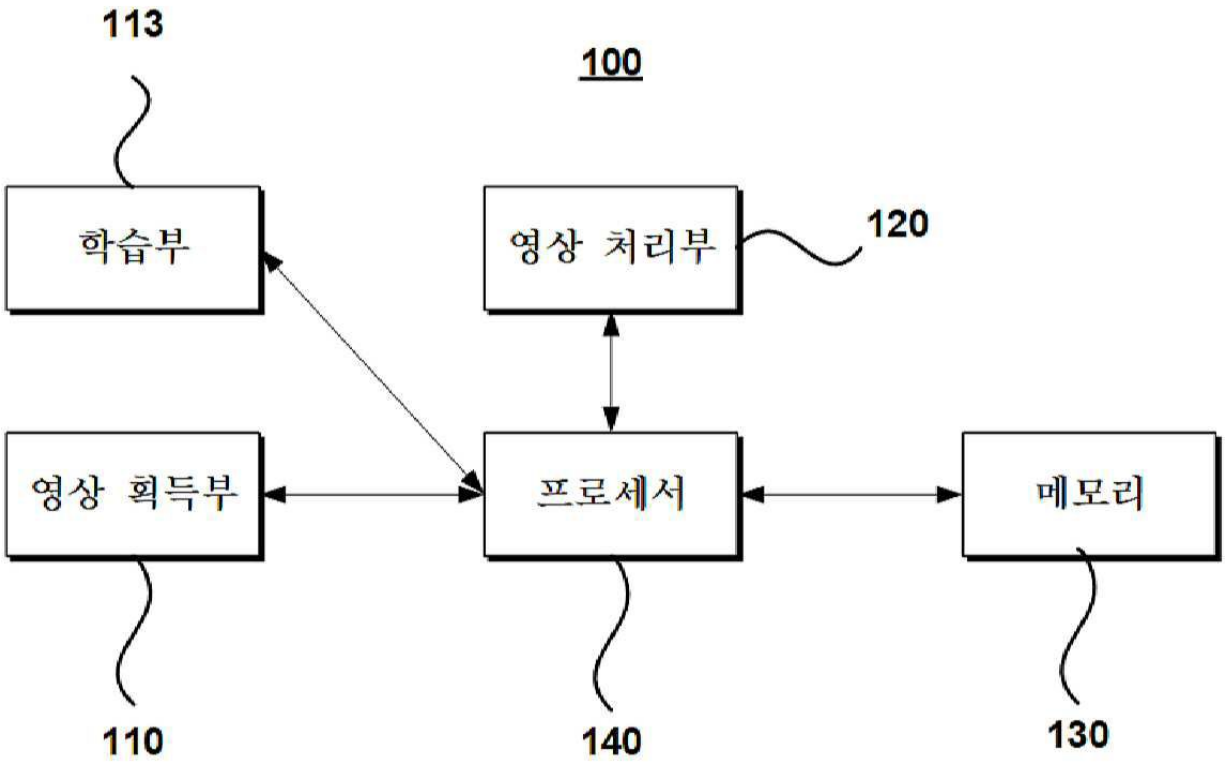
[0096] 상술한 하드웨어 장치는 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될

수 있으며, 그 역도 마찬가지이다.

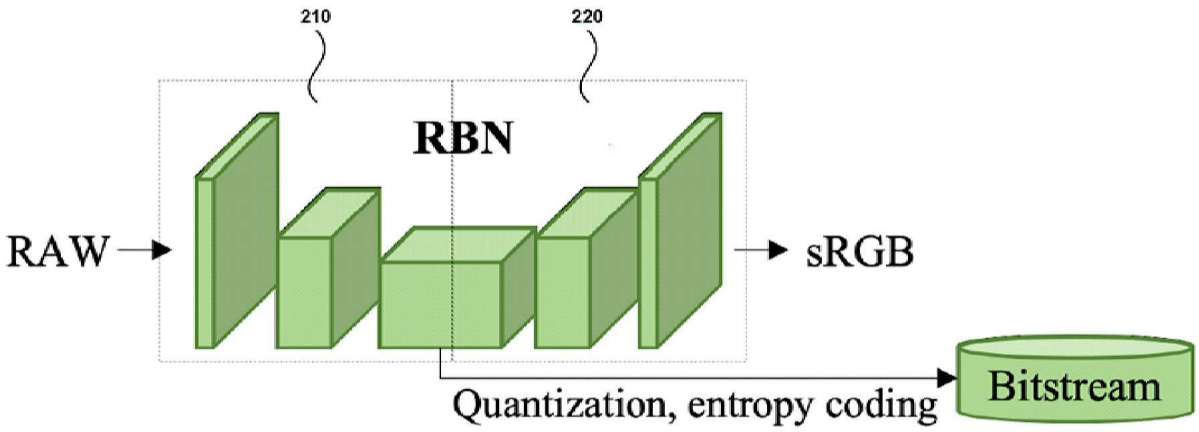
[0097] 이제까지 본 발명에 대하여 그 실시 예들을 중심으로 살펴보았다. 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시 예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

도면

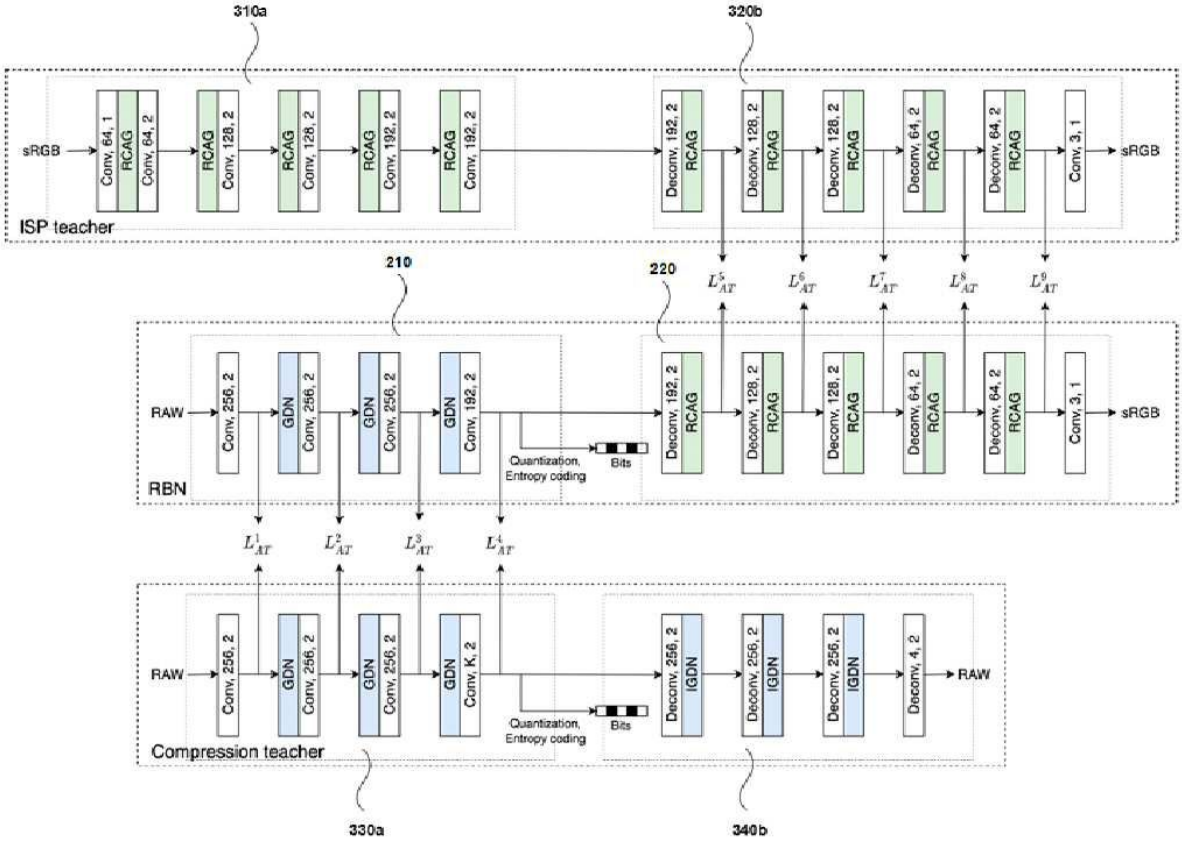
도면1



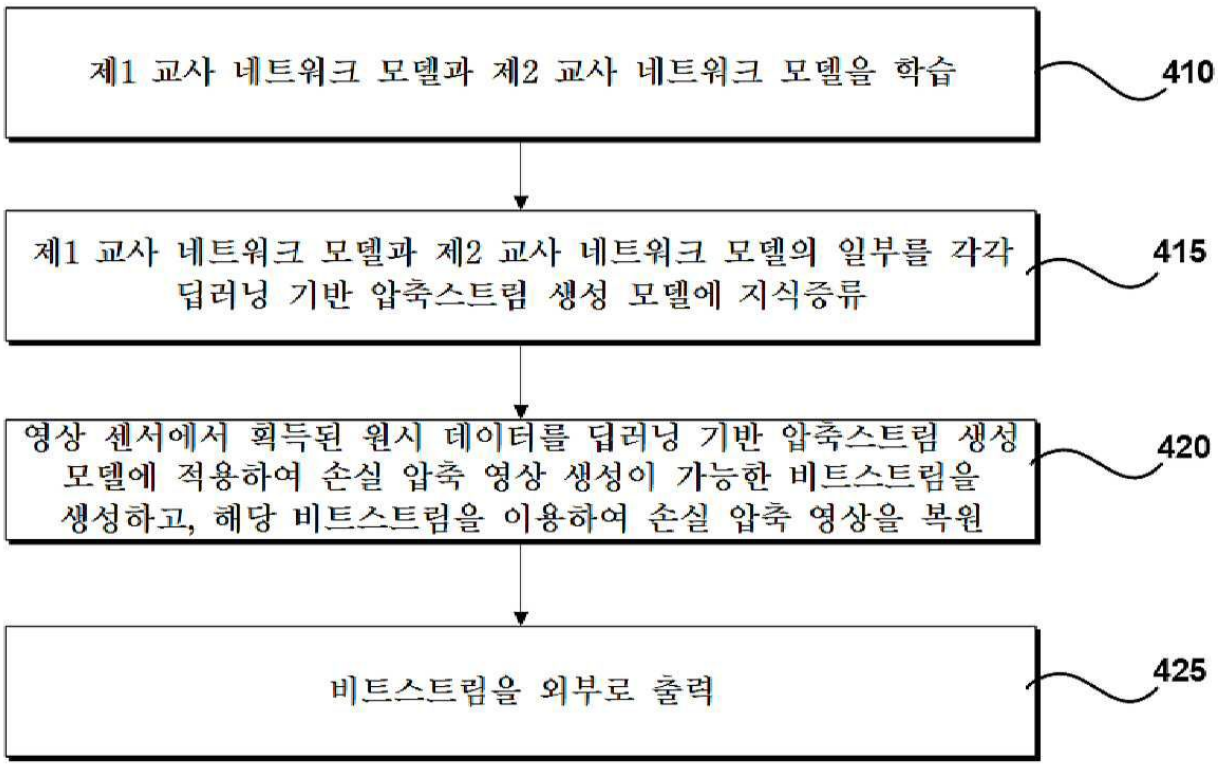
도면2



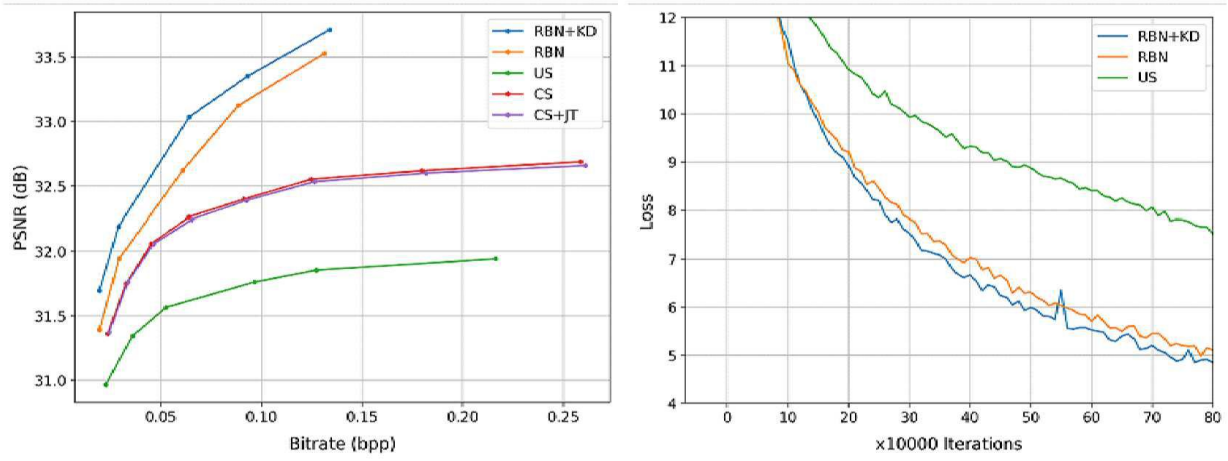
도면3


310a
320b
sRGB
Conv, 64, 1
RCAG
Conv, 64, 2
RCAG
Conv, 128, 2
RCAG
Conv, 128, 2
RCAG
Conv, 192, 2
RCAG
Conv, 192, 2
Deconv, 192, 2
RCAG
Deconv, 128, 2
RCAG
Deconv, 128, 2
RCAG
Deconv, 64, 2
RCAG
Deconv, 64, 2
RCAG
Conv, 3, 1
sRGB
ISP teacher
210
220
RAW
Conv, 256, 2
GDN
Conv, 256, 2
GDN
Conv, 256, 2
GDN
Conv, 192, 2
Quantization, Entropy coding
Bits
RBN
Conv, K, 2
Deconv, 256, 2
IGDN
Deconv, 4, 2
Compression teacher
330a
340b

도면4

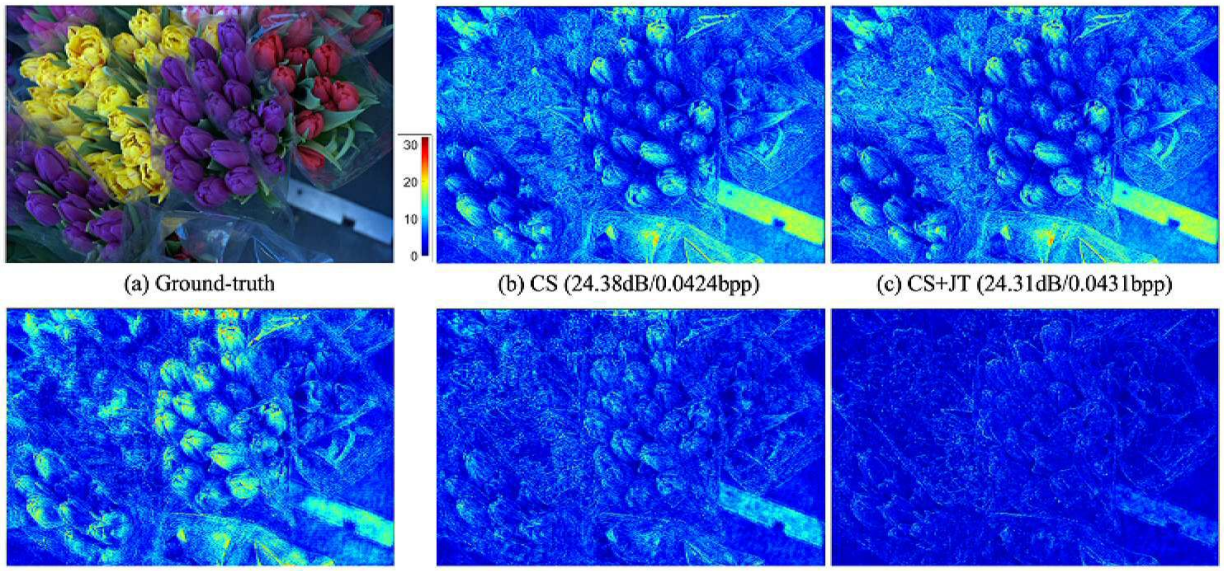

제1 교사 네트워크 모델과 제2 교사 네트워크 모델을 학습
410
제1 교사 네트워크 모델과 제2 교사 네트워크 모델의 일부를 각각 딥러닝 기반 압축스트림 생성 모델에 지식증류
415
영상 센서에서 획득된 원시 데이터를 딥러닝 기반 압축스트림 생성 모델에 적용하여 손실 압축 영상 생성이 가능한 비트스트림을 생성하고, 해당 비트스트림을 이용하여 손실 압축 영상을 복원
420
비트스트림을 외부로 출력
425

도면5



(a) (b)

도면6



(a) Ground-truth (b) CS (24.38dB/0.0424bpp) (c) CS+JT (24.31dB/0.0431bpp)

(d) US (25.56dB/0.0493bpp) (e) RBN (27.80dB/0.0417bpp) (f) RBN+KD (32.51dB/0.0400bpp)

도면7



(a) Ground-truth (b) CS (32.66dB/0.0674bpp) (c) CS+JT (32.59dB/0.0679bpp) (d) US (33.55dB/0.0844bpp) (d) RBN (35.17dB/0.0578bpp) (e) RBN+KD (37.84dB/0.0617bpp)